Open access • Journal Article • DOI:10.3982/ECTA15867

Identification With Additively Separable Heterogeneity — Source link

Roy Allen, John Rehbeck

Institutions: University of Western Ontario, Ohio State University

Published on: 01 May 2019 - Econometrica (John Wiley & Sons, Ltd)

Topics: Unobservable and Discrete choice

Econometrica Supplementary Material

SUPPLEMENT TO "IDENTIFICATION WITH ADDITIVELY SEPARABLE HETEROGENEITY"

(*Econometrica*, Vol. 87, No. 3, May 2019, 1021–1054)

ROY ALLEN
Department of Economics, University of Western Ontario

JOHN REHBECK
Department of Economics, The Ohio State University

This supplement provides additional proofs and results for the authors' paper "Identification with Additively Separable Heterogeneity."

S.1. PRELIMINARIES IN CONVEX ANALYSIS

DEFINITION S.1.1—Subdifferential: Let $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The subdifferential of f at z is the set

$$\partial f(z) = \{w \in \mathbb{R}^K : \forall \tilde{z} \in \mathbb{R}^K, f(\tilde{z}) - f(z) \geq w'(\tilde{z} - z)\}.$$

An element of $\partial f(z)$ is called a subgradient at z.

DEFINITION S.1.2—Convex Conjugate: Let $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$. Then the convex conjugate of f is denoted

$$f^*(w) = \sup_{z \in \mathbb{R}^K} \{z'w - f(z)\}.$$

The function f^* is convex (regardless of whether f is convex) as discussed in Rockafellar (1970, p. 104).

LEMMA S.1.1—Rockafellar (1970, Theorem 23.5): *Let $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be a convex function such that $f(y) < \infty$ for some y. Then the following are equivalent*:
(i) $w \in \partial f(\tilde{z})$.
(ii) *$z'w - f(z)$ attains its supremum in z at $z = \tilde{z}$.*
If, in addition, f is lower semi-continuous,[1] *then the following conditions are also equivalent to the ones above*:
(iii) $\tilde{z} \in \partial f^*(\tilde{w})$.
(iv) *$w'\tilde{z} - f^*(w)$ attains its supremum in w at $w = \tilde{w}$.*

LEMMA S.1.2—Rockafellar (1970, Theorem 25.1): *Let $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be convex and assume $f(z) < \infty$. Then f is differentiable at z if and only if $\partial f(z)$ is a singleton.*

Roy Allen: rallen46@uwo.ca
John Rehbeck: rehbeck.7@osu.edu

[1]That is, for each α, $\{y \in \mathbb{R}^K \mid f(y) > \alpha\}$ is open.

 https://doi.org/10.3982/ECTA15867

S.2. MEASURABILITY

In this section, we provide sufficient conditions for a measurable selector to exist in PUM.

Let $\varepsilon : \Omega_\varepsilon \to E$ be a random variable defined from the probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$ to the measure space $(E, \mathcal{E})$. Let $X : \Omega_X \to \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_K}$ be a random variable from the probability space $(\Omega_X, \mathcal{F}_X, P_X)$ to the measurable space $(\prod_{k=1}^K \mathbb{R}^{d_k}, \mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}})$, where $\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}}$ denotes the Borel σ-algebra induced by the Euclidean metric in $\prod_{k=1}^K \mathbb{R}^{d_k}$. Note that the projection of X in the kth component is a random variable to $(\mathbb{R}^{d_k}, \mathcal{B}_{\mathbb{R}^{d_k}})$.

ASSUMPTION S.2.1: *We assume the disturbance function D, utility indices $\vec{u}$, and the budget B satisfy the following*:

(i) $D : \mathbb{R}^K \times E \to \mathbb{R} \cup \{-\infty\}$. *For each* $y \in \mathbb{R}^K$, $D(y, \cdot)$ *is* $\mathcal{E}$*-measurable. For each* $\varepsilon \in \mathcal{E}$, $D(\cdot, \varepsilon)$ *is continuous.*

(ii) *For all* $k = 1, \dots, K$, $u_k : \mathbb{R}^{d_k} \to \mathbb{R}$ *is* $\mathcal{B}_{\mathbb{R}^{d_k}}$*-measurable.*

(iii) $B \subseteq \mathbb{R}^K$ *is a nonempty, compact, and convex budget.*

LEMMA S.2.1: *If Assumption S.2.1 holds and X and ε are independent, there exists a measurable selector*

$$Y^* \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^K y_k u_k(X_k) + D(y, \varepsilon).$$

PROOF: From Assumption S.2.1(i) and Stinchcombe and White (1992, Lemma 2.15), D is $\mathcal{B}_{\mathbb{R}^K} \otimes \mathcal{E}$-measurable. From Assumption S.2.1(ii), we have that $u_k(X_k)$ is $\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}}$-measurable since it is a composition of a measurable function and the continuous projection of X to X_k. Therefore, each $g_k(y, X) = y_k u_k(X_k)$ is continuous in $\mathbb{R}^K$ for each X and $\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}}$-measurable for each $y \in \mathbb{R}^k$. By Stinchcombe and White (1992, Lemma 2.15), g_k is $\mathcal{B}_{\mathbb{R}^K} \otimes \mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}}$-measurable.

Since $D(y, \varepsilon)$ does not depend on X and X and ε are independent, we may extend D to be $\mathcal{B}_{\mathbb{R}^K} \otimes (\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}} \otimes \mathcal{E})$-measurable.[2] Similarly, each $g_k(y, X)$ does not depend on ε and X and ε are independent, so we may extend g_k to be $\mathcal{B}_{\mathbb{R}^K} \otimes (\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}} \otimes \mathcal{E})$-measurable. Therefore, $U(\cdot; X, \varepsilon) = \sum_{k=1}^K y_k u_k(X_k) + D(y, \varepsilon)$ is $\mathcal{B}_{\mathbb{R}^K} \otimes (\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}} \otimes \mathcal{E})$-measurable since it is the sum of measurable functions. Moreover, for all $(X, \varepsilon) \in \prod_{k=1}^K \mathbb{R}^{d_k} \times E$, $U(\cdot; X, \varepsilon) : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty, -\infty\}$ is continuous.

Last, from Assumption S.2.1(iii), the assumptions of Corollary 2.21 and Lemma 2.22 in Stinchcombe and White (1992) are satisfied, so for any probability measure P on $((\prod_{k=1}^K \mathbb{R}^{d_k}, E), (\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}} \otimes \mathcal{E}))$, there is a $(\mathcal{B}_{\prod_{k=1}^K \mathbb{R}^{d_k}} \otimes \mathcal{E})^P$-measurable everywhere selection[3] Y^* that satisfies $Y^* \in \operatorname{argmax}_{y \in B} U(y; X, \varepsilon)$. Note that $\operatorname{argmax}_{y \in B} U(y; X, \varepsilon)$ is nonempty and compact for all $(X, \varepsilon) \in \prod_{k=1}^K \mathbb{R}^{d_k} \times E$. *Q.E.D.*

S.3. IDENTIFICATION FOR COMMON REGRESSORS WITH WEAKER SUPPORT CONDITIONS

This section describes how the arguments of Section 3.2 may be adapted to handle violations of Assumption 3. First, suppose that there is a subset $\mathcal{W} \subseteq \operatorname{supp}(W)$ such that

[2] We suppress dependence of D on X for convenience.

[3] $\mathcal{F}^P$ is the completion of a σ-field $\mathcal{F}$ with respect to the probability measure P.

for each $w \in \mathcal{W}$, $\vec{u}(z, w)$ is identified up to location and scale for all $z \in \operatorname{supp}(Z \mid W = w)$. Theorem 2 provides conditions for this by conditioning. However, Theorem 2 does not provide conditions under which these functions $\vec{u}(\cdot, w)$, indexed by w, are identified up to a common scale. We will first describe how to identify all functions $\vec{u}(\cdot, w)$ up to the location and scale set by $\vec{u}(\cdot, w^*)$ for a specific $w^* \in \mathcal{W}$. Suppose that w and w^* have overlap in the sense that $\operatorname{int}(\vec{u}(\operatorname{supp}(Z \mid W = w), w) \cap \vec{u}(\operatorname{supp}(Z \mid W = w^*), w^*)) \neq \emptyset$, where for a set S, $\operatorname{int}(S)$ denotes the interior. Because the representative agent problem has a unique solution, this implies that there exist values $(z', w')', (z^{*\prime}, w^{*\prime})' \in \operatorname{supp}(Z, W)$ such that $\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = z^*, W = w^*]$. If $\overline{D}_{\overline{B}}$ is differentiable at this common point, then the level of $\vec{u}(z, w)$ may be calibrated to equal that of $\vec{u}(z^*, w^*)$ by Lemma 3. This argument identifies the function $\vec{u}(\cdot, w)$ up to the same *location* as that of $\vec{u}(\cdot, w^*)$. Similarly, by varying z^*, the *scale* of $\vec{u}(\cdot, w)$ may also be calibrated relative to that of $\vec{u}(\cdot, w^*)$. This is possible because we assumed the interior of the intersection is nonempty, $\operatorname{int}(\vec{u}(\operatorname{supp}(Z \mid W = w), w) \cap \vec{u}(\operatorname{supp}(Z \mid W = w^*), w^*)) \neq \emptyset$.

Now, it may happen that w and w^* do not have the required overlap. To handle this, define the mapping

$$E(\tilde{w}) = \Big\{ w \in \mathcal{W} \mid \operatorname{int}\Big(\vec{u}\big(\operatorname{supp}(Z \mid W = w), w\big) \cap \vec{u}\big(\operatorname{supp}(Z \mid W = \tilde{w}), \tilde{w}\big)\Big) \neq \emptyset \Big\}.$$

By the previous arguments, the functions $\vec{u}(\cdot, w)$ are identified up to a common location and scale for each $w \in E(w^*)$. Note for $n \geq 2$, we define $E^n(\tilde{w}) = E(E^{n-1}(\tilde{w}))$. Now, if $\bigcup_{m=1}^{\infty} E^m(w^*) = \mathcal{W}$ and if $\bigcup_{w \in \mathcal{W}} \vec{u}(\operatorname{supp}(Z \mid W = w), w) = \vec{u}(\operatorname{supp}(Z, W))$, we may identify $\vec{u}(z, w)$ for each $(z', w')' \in \operatorname{supp}(Z, W)$ under the assumption that $\overline{D}_{\overline{B}}$ is differentiable. The argument is analogous to the proof of Theorem 3. Thus, we have proven the following result.

PROPOSITION S.3.1: *Assume there is a subset $\mathcal{W} \subseteq \operatorname{supp}(W)$ such that, for each $w \in \mathcal{W}$, $\vec{u}(z, w)$ is identified up to location and scale for all $z \in \operatorname{supp}(Z \mid W = w)$, $\bigcup_{m=1}^{\infty} E^m(w^*) = \mathcal{W}$ for some $w^* \in \mathcal{W}$, $\bigcup_{w \in \mathcal{W}} \vec{u}(\operatorname{supp}(Z \mid W = w), w) = \vec{u}(\operatorname{supp}(Z, W))$, and $\overline{D}_{\overline{B}}$ is differentiable. Then $\vec{u}(z, w)$ is identified for each $z, w \in \operatorname{supp}(Z, W)$ up to a common location and scale.*

S.4. MAXIMUM SCORE

We demonstrate how a generalization of the maximum score inequalities (Manski (1975), Matzkin (1993), Goeree, Holt, and Palfrey (2005), Fox (2007)) may be obtained with a symmetry condition on the aggregate disturbance function $\overline{D}$. These inequalities provide identifying information even for a fixed value of regressors.

PROPOSITION S.4.1: *Let the assumptions of Theorem* 1 *hold and assume* $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y(x, \varepsilon)]$ *for some* $x \in \operatorname{supp}(X)$. *Assume for every permutation π of the good indices* $\{1, \dots, K\}$, $(y_1, \dots, y_K) \in \overline{B}$ *implies that the permutation is in the budget*, $(y_{\pi(1)}, \dots, y_{\pi(K)}) \in \overline{B}$, *and the disturbance is the same for the permutation*, $\overline{D}(y_1, \dots, y_K) = \overline{D}(y_{\pi(1)}, \dots, y_{\pi(K)})$. *Then*

$$\mathbb{E}[Y_k \mid X = x] > \mathbb{E}[Y_\ell \mid X = x] \quad \Longrightarrow \quad u_k(x_k) \geq u_\ell(x_\ell)$$

and

$$u_k(x_k) > u_\ell(x_\ell) \quad \Longrightarrow \quad \mathbb{E}[Y_k \mid X = x] \geq \mathbb{E}[Y_\ell \mid X = x].$$

PROOF: For notational simplicity, we write permutations in terms of matrices rather than functions permutating the arguments. To that end, let Π be a permutation matrix, that is, a matrix whose rows and columns each sum to 1 and whose entries each consist of 0 or 1. Because the conditional mean is a maximizer and $y \in \overline{B} \implies \Pi y \in \overline{B}$, we have

$$E[Y \mid X = x]'\vec{u}(x) + \overline{D}\big(\mathbb{E}[Y \mid X = x]\big) \geq E[\Pi Y \mid X = x]'\vec{u}(x) + \overline{D}\big(\mathbb{E}[\Pi Y \mid X = x]\big).$$

Because $\overline{D}$ is permutation symmetric, both terms involving $\overline{D}$ are equal and we obtain

$$\big(E[Y \mid X = x] - E[\Pi Y \mid X = x]\big)'\vec{u}(x) \geq 0.$$

Letting Π be a permutation matrix that permutes the kth and ℓth components of each vector, we obtain

$$\big(\mathbb{E}[Y_k \mid X = x] - \mathbb{E}[Y_\ell \mid X = x]\big)\big(u_k(x_k) - u_\ell(x_\ell)\big) \geq 0,$$

and the result follows. *Q.E.D.*

These inequalities hold even if $\text{supp}(X) = x$ is a singleton, in which case X and ε are trivially independent. Thus, independence between X and ε is not essential for these inequalities. Indeed, the proposition holds as well if x enters $\overline{D}$, provided $\overline{D}$ is symmetric to permutations in y for each fixed x.

In contrast with our main identification results for utility indices presented in Section 3, these identifying inequalities are purely ordinal, that is, they cannot distinguish between $\vec{u} = (u_1, \dots, u_K)$ and the composition $(g \circ u_1, \dots, g \circ u_K)$ for a strictly increasing function g. We conjecture that if $\text{supp}(X) = x$, that is, there is no variation in X, these inequalities are sharp.

Specialized to ARUM, it is easy to see that $\overline{D}$ is symmetric to permutations whenever the distribution of ε is exchangeable conditional on $X = x$ (i.e., the joint distribution of ε is invariant to relabeling indices). In the binary choice case (Manski (1975)), using the fact that probabilities sum to 1, the implications above specialize to

$$\mathbb{E}[Y_1 \mid X = x] > \frac{1}{2} \quad \implies \quad u_1(x_1) \geq u_2(x_2)$$

and

$$u_1(x_1) > u_2(x_2) \quad \implies \quad \mathbb{E}[Y_1 \mid X = x] \geq \frac{1}{2}.$$

In the bundles model, $\overline{D}$ is symmetric to permutations if, conditional on $\varepsilon_{1,1}$, the joint distribution of $(\varepsilon_{1,0}, \varepsilon_{0,1})$ is exchangeable.

S.5. IDENTIFICATION OF UTILITY INDICES FOR "NONSTANDARD" CASES

As discussed previously, Theorem 2 rules out some examples of interest. We now provide weaker conditions under which utility indices are identified. Instead of assuming mixed partial derivatives of V are everywhere nonzero, we assume mixed partials are nonzero at a "rich" set of points. The building block for this extension is the observation that Proposition 1 provides a constructive identification formula using only local assumptions concerning two goods at a time.

Recall that if we identify many ratios of partial derivatives of $\vec{u}$, then we can identify $\vec{u}$ itself by two different approaches. The first, which is feasible given the assumptions of

Theorem 2, allows us to integrate these derivatives and obtain constructive identification. The second approach, taken in this section, is to use the mean value theorem to obtain nonconstructive results. Recall that by the mean value theorem, a differentiable function is uniquely determined by its partial derivatives up to location. Thus, we only need to identify ratios of all partial derivatives of $\vec{u}$. We describe how to do this by multiplying ratios that are directly identified by Proposition 1.

For an example, suppose

$$\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=x_\ell^*}\bigg/\frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x_k^*} \quad \text{and} \quad \frac{\partial u_j(x_j)}{\partial x_{j,r}}\bigg|_{x_j=\tilde{x}_j}\bigg/\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=\tilde{x}_\ell}$$

are identified from data (via Proposition 1). If x^* and $\tilde{x}$ agree for the regressors of good ℓ, that is, $x_\ell^* = \tilde{x}_\ell$, we can multiply these derivative ratios to identify

$$\frac{\partial u_j(x_j)}{\partial x_{j,r}}\bigg|_{x_j=\tilde{x}_j}\bigg/\frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x_k^*}. \tag{S.1}$$

Note that this derivative ratio can be identified *even if* $j = k$ and $\tilde{x}_j \neq x_k^*$; that is, we can identify derivative ratios for the same good at different values of regressors. This shows that if ratios of derivatives are identified and overlap in a specific sense, then we can multiply these ratios to identify new ratios. We present a result that handles sequences of derivative ratios of arbitrary finite length, so we introduce some more notation.

DEFINITION S.5.1—Paths: There is a path from the point $\tilde{a} := \frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*}$ to $\tilde{b} := \frac{\partial u_j(x_j)}{\partial x_{j,r}}|_{x_j=\tilde{x}_j}$ if $\tilde{a}$ is nonzero and there is a sequence of partial derivatives beginning at $\tilde{a}$ and ending at $\tilde{b}$ such that each adjacent element is paired, and these pairs are strict except possibly between the final two elements of the sequence.

In order for there to be a path between partial derivatives, several conditions must hold. If we have multiple goods ($K \geq 2$), then there can be a path between partial derivatives at different values x_k^* and $\tilde{x}_j$. However, if there is only a single good, then the derivative ratios of different regressors must be evaluated at the same point, that is, $x_k^* = \tilde{x}_j$ and $j = k$. The function V must be twice continuously differentiable at a sufficiently rich set of points. Importantly, it is *not* necessary that all mixed partials be nonzero or that V be twice continuously differentiable everywhere. This condition is in the spirit of the connected substitutes condition of Berry, Gandhi, and Haile (2013). This condition differs from connected substitutes since we look over paths derived from either complementarity or substitutability. We also differ since the path condition only involves two goods at a time, while the connected substitutes condition places restrictions on all subsets of goods.

THEOREM S.5.1: *Let Assumption* 2 *hold and assume* $x_{k,q}$ *and* $x_{j,r}$ *are regressors specific to* k *and* j, *respectively. If there is a path from the point* $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*}$ *to* $\frac{\partial u_j(x_j)}{\partial x_{j,r}}|_{x_j=\tilde{x}_j}$, *then*

$$\frac{\partial u_j(x_j)}{\partial x_{j,r}}\bigg|_{x_j=\tilde{x}_j}\bigg/\frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x_k^*} \tag{S.2}$$

is identified.

PROOF OF THEOREM S.5.1: Let $\frac{\partial u_{\ell_1}(x_{\ell_1})}{\partial x_{\ell_1,p_1}}|_{x_{\ell_1}=x^1_{\ell_1}}, \dots, \frac{\partial u_{\ell_M}(x_{\ell_M})}{\partial x_{\ell_M,p_M}}|_{x_{\ell_M}=x^M_{\ell_M}}$ be a finite sequence as in Definition S.5.1. For each $m = 2, \dots, M$, let

$$S_{m-1,m} = \frac{\partial u_{\ell_m}(x_{\ell_m})}{\partial x_{\ell_m,p_m}}\bigg|_{x_{\ell_m}=x^m_{\ell_m}} \bigg/ \frac{\partial u_{\ell_{m-1}}(x_{\ell_{m-1}})}{\partial x_{\ell_{m-1},p_{m-1}}}\bigg|_{x_{\ell_{m-1}}=x^{m-1}_{\ell_{m-1}}}. \tag{S.3}$$

This ratio is identified due to Proposition 1. This follows because for $m < M$, the numerator and denominator are strictly paired. For $m = M$, the numerator and denominator are paired and the denominator is nonzero.

By construction,

$$\prod_{m=2}^{M} S_{m-1,m} = \frac{\partial u_j(x_j)}{\partial x_{j,r}}\bigg|_{x_j=\tilde{x}_j} \bigg/ \frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x^*_k}$$

is identified since all intermediate terms cancel out. This is valid because there is never any division by zero. *Q.E.D.*

The following corollary relaxes assumptions in Theorem 2.

COROLLARY S.5.1: *Let Assumption* 2 *hold and assume all regressors are good-specific. Assume there is a tuple* (k, q, x^*_k) *such that* $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k}$ *has a path to* $\frac{\partial u_j(x_j)}{\partial x_{j,r}}|_{x_j=\tilde{x}_j}$ *for any* j, r, *and* $\tilde{x}_j \in \mathbb{R}^{d_j}$. *Then* $\vec{u}$ *is identified under the following normalization*:

(i) (*Scale*) $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k} \in \{-1, 1\}$.

(ii) (*Location*) $u_\ell(0_{d_k}) = 0$ *for each* $\ell = 1, \dots, K$, *where* 0_{d_k} *denotes a* d_k*-dimensional vector of zeros*.

PROOF OF COROLLARY S.5.1: First, we identify the sign of $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k}$. Under Assumption 2, it can be shown that

$$\begin{aligned} &\mathbb{E}[Y \mid X = x] \neq \mathbb{E}[Y \mid X = \tilde{x}] \\ &\iff \quad \big(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}]\big)'\big(\vec{u}(x) - \vec{u}(\tilde{x})\big) > 0. \end{aligned} \tag{S.4}$$

(This is a straightforward extension of Lemma S.7.1.) From the assumptions of the corollary, there is some $x^* \in \operatorname{supp}(X)$ that has x^*_k as its kth row. Moreover, there must be some ℓ such that

$$\partial_{k,\ell} V(\overline{u})|_{\overline{u}=\vec{u}(x^*)} \neq 0.$$

This follows from the definition of a path. Since $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k} \neq 0$, we see that, for sufficiently small changes in $x_{k,q}$, there must be a change in $\mathbb{E}[Y_\ell \mid X = x]$. From (S.4), this implies that there must be a change in $\mathbb{E}[Y_k \mid X = x]$ as well. Again using (S.4), we determine the sign of $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k}$ depending on whether $\mathbb{E}[Y_k \mid X = x]$ is locally increasing or decreasing with respect to $x_{k,q}$.

Normalizing $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x^*_k}$ to 1 or -1 depending on its sign, we identify all partial derivatives using Theorem S.5.1. Recall that from the mean value theorem, two functions that share partial derivatives can differ by at most an additive constant. Given the location normalization, $\vec{u}$ is identified. *Q.E.D.*

We provide an example where the path condition holds even though $\mathbb{E}[Y \mid X = x]$ may lie on the boundary of $\overline{B}$ and may not even be continuous. This illustrates how identification in Corollary S.5.1 is established using restrictions on pairs of partial derivatives. This is in contrast with Theorem 2, which places global restrictions that rule out boundary behavior and discontinuities as in the following example.

REMARK S.5.1—Boundaries and Discontinuities: For illustration of how we can handle boundary issues, let $\overline{B}$ be the probability simplex, $K > 2$, and let $\overline{D}$ be given by

$$\overline{D}(y) = \begin{cases} -\sum_{k=1}^{K} p_k \ln p_k, & \text{if } p_k \neq 0 \text{ for at most 2 distinct } k, \\ -\infty, & \text{otherwise.} \end{cases}$$

We set $0 \ln 0$ to 0. For simplicity, suppose x_k is scalar. This choice of $\overline{D}$ ensures exactly two goods will be chosen with positive probability, and requires that they be the ones with the highest values of the indices $u_k(x_k)$ (assuming the two highest values are unique). Suppose the second highest value of $u_k(x_k)$ is unique and let $k(1)$ and $k(2)$ attain the highest values of u_k.[4] Then

$$\mathbb{E}[Y_{k(1)} \mid X = x] = \frac{e^{u_{k(1)}(x_{k(1)})}}{e^{u_{k(1)}(x_{k(1)})} + e^{u_{k}(2)(x_{k(2)})}},$$

$$\mathbb{E}[Y_{k(2)} \mid X = x] = \frac{e^{u_{k(2)}(x_{k(2)})}}{e^{u_{k(1)}(x_{k(1)})} + e^{u_{k(2)}(x_{k(2)})}}.$$

If $\vec{u}$ is differentiable, sufficient conditions for Corollary S.5.1 are fairly mild. One sufficient condition is that X has full support, all partial derivatives of $\vec{u}$ are everywhere nonzero, and $\vec{u}(\text{supp}(X)) = \mathbb{R}^K$.

S.6. INJECTIVITY ON THE SIMPLEX

For these results, we assume $\overline{B}$ is the probability simplex,

$$\overline{B} = \left\{ y \in \mathbb{R}^K \;\middle|\; \sum_{k=1}^{K} y_k = 1, \, y_k \geq 0 \text{ for } k = 1, \ldots, K \right\}.$$

In order to obtain an injectivity result, we need to restrict the set of possible values of the vector $\vec{v}$. This is because, for fixed $\overline{D}$,

$$\operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k v_k + \overline{D}(y)$$

is the same set with $\vec{v}$ replaced by $\vec{v} + \vec{c}$, where $\vec{c}$ is a constant vector. We restrict the parameter space for $\vec{v}$ with the following normalization for its first component:

$$\mathcal{V} = \left\{ \vec{v} \in \mathbb{R}^K \mid v_1 = 0 \right\}.$$

[4] These implicitly depend on x.

We are now interested in when

$$\rho_{\mathcal{V}}^{-1}(y^*) = \left\{\vec{v} \in \mathcal{V} \;\middle|\; y^* \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k v_k + \overline{D}(y)\right\}$$

is a singleton.

ASSUMPTION S.6.1: *$\overline{D} : \overline{B} \to \mathbb{R} \cup \{-\infty\}$ is a concave function. Moreover, $\{y \in \overline{B} \mid \overline{D}(y) > -\infty\}$ has nonempty interior when viewed as a subset of $\overline{B}$.*

A nonempty interior will be needed because we invoke differentiability of $\overline{D}$. Let

$$T = \left\{y \in \operatorname{ri}(\overline{B}) \;\middle|\; y \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k v_k + \overline{D}(y) \text{ for some } \vec{v} \in \mathcal{V}\right\},$$

where $\operatorname{ri}(\overline{B})$ denotes the relative interior of $\overline{B}$.[5]

PROPOSITION S.6.1: *Let $y^* \in T$, let $K \geq 2$, let $\overline{B}$ be the probability simplex, and let Assumption S.6.1 hold. Then the following are equivalent:*
(i) *$\rho_{\mathcal{V}}^{-1}(y^*)$ is a singleton.*
(ii) *$\overline{D}$ is Fréchet differentiable at y^*.*

PROOF: We can prove this from Lemmas S.1.1 and S.1.2 with a change of variables. The basic idea will be that over the probability simplex, y_1 is uniquely determined by $(y_2, \dots, y_K)$. Using the normalization defining $\mathcal{V}$, we may convert the problem from a K-dimensional problem to a $K-1$-dimensional problem and then invoke Lemmas S.1.1 and S.1.2.

We define a new function,

$$\tilde{D}(y_2, \dots, y_K) = \begin{cases} \overline{D}\left(\left(1 - \sum_{k=2}^{K} y_k\right), y_2, \dots, y_K\right), & \text{if } \sum_{k=2}^{K} y_k \leq 1, y_k \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

This function removes y_1 by using the budget constraint. The function $\tilde{D}$ is concave since $\overline{D}$ is concave and $\overline{B}$ is convex.

Now define the multi-valued mapping $\rho_{\mathcal{V}}$ that maps points in $\mathbb{R}^{K-1}$ to subsets of $\mathbb{R}^K$ by

$$\rho_{\mathcal{V}}(v_2, \dots, v_K) = \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=2}^{K} y_k v_k + \overline{D}\left(\left(1 - \sum_{k=2}^{K} y_k\right), y_2, \dots, y_K\right). \tag{S.5}$$

[5]The relative interior of $\overline{B}$ is the set

$$\operatorname{ri}(\overline{B}) = \left\{y \in \mathbb{R}^K \;\middle|\; \sum_{k=1}^{K} y_k = 1, y_k > 0 \text{ for } k = 1, \dots, K\right\}.$$

Note that the choice of the first component (y_1) now enters trivially. Similarly define ρ from $\mathbb{R}^{K-1}$ to subsets of $\mathbb{R}^{K-1}$ by

$$\rho(v_2, \dots, v_K) = \underset{y_2,\dots,y_K \in \mathbb{R}^{K-1}}{\operatorname{argmax}} \sum_{k=2}^{K} y_k \vec{v}_k + \tilde{D}(y_2, \dots, y_K). \tag{S.6}$$

Over the probability simplex, we may put $\rho_{\mathcal{V}}(\vec{v})$ and $\rho(\vec{v})$ in one-to-one correspondence by the projection mapping $\pi(y_1, \dots, y_K) = (y_2, \dots, y_K)$.

Finally, note Fréchet differentiability of $\overline{D}$ at y^* is equivalent to differentiability of $\tilde{D} : \mathbb{R}^{K-1} \to \mathbb{R} \cup \{+\infty\}$ at $\pi(y^*)$. Invoking Lemmas S.1.1 and S.1.2, $\rho_{\mathcal{V}}^{-1}(y^*)$ is a singleton if and only if $\overline{D}$ is Fréchet differentiable at y^*. *Q.E.D.*

S.7. PARTIAL IDENTIFICATION OF UTILITY INDICES

Our sufficient conditions for identification of utility indices may fail. Our conditions do not apply if all regressors are discrete, sufficient substitution/complementarity does not exist, or if $\mathbb{E}[Y \mid X = x]$ is not suitably differentiable. We provide a complete characterization of the identifying power of the model for utility indices.

Now, we allow the possibility that there are multiple functions $\vec{u}$ that are consistent with the restrictions of the model. First, let $\mathcal{D}$ denote the set of admissible disturbances, which are functions $\tilde{D} : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ that are concave, upper semi-continuous, and finite at some $y \in \overline{B}$. We denote the *identified set* for $\vec{u}$ as

$$\mathcal{U}_{\mathrm{ID}} = \Bigg\{ \vec{u} \in \mathcal{U} \mid \exists \tilde{D} \in \mathcal{D} \text{ s.t. } \forall x \in \operatorname{supp}(X),$$
$$\mathbb{E}[Y \mid X = x] \in \underset{y \in \overline{B}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k u_k(x_k) + \tilde{D}(y) \Bigg\}.$$

Note that in this section, we do not assume the conditional mean is the unique maximizer.

The set $\mathcal{U}$ is the parameter space for $\vec{u}$. We assume $\mathcal{U}$ consists of real-valued functions. Any further restrictions on this set shrink $\mathcal{U}_{\mathrm{ID}}$. For example, it could be a parametric class of functions.

The following lemma is a convenient restatement of restrictions of the optimizing model.

LEMMA S.7.1: *If $\vec{u} \in \mathcal{U}_{\mathrm{ID}}$, then there is some $\tilde{D} \in \mathcal{D}$ such that for every $x, \tilde{x} \in \operatorname{supp}(X)$,*

$$\begin{aligned}\big(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}]\big)' \vec{u}(x) &\geq \tilde{D}\big(\mathbb{E}[Y \mid X = \tilde{x}]\big) - \tilde{D}\big(\mathbb{E}[Y \mid X = x]\big) \\ &\geq \big(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}]\big)' \vec{u}(\tilde{x}).\end{aligned}$$

Moreover, $\tilde{D}(\mathbb{E}[Y \mid X = \tilde{x}])$ and $\tilde{D}(\mathbb{E}[Y \mid X = x])$ are finite.

PROOF: We use necessary conditions for optimality. If $\vec{u} \in \mathcal{U}_{\mathrm{ID}}$, then, for some $\tilde{D} \in \mathcal{D}$, we must have

$$\mathbb{E}[Y \mid X = x]' \vec{u}(x) + \tilde{D}\big(\mathbb{E}[Y \mid X = x]\big) \geq \mathbb{E}[Y \mid X = \tilde{x}]' \vec{u}(x) + \tilde{D}\big(\mathbb{E}[Y \mid X = \tilde{x}]\big),$$
$$\mathbb{E}[Y \mid X = \tilde{x}]' \vec{u}(\tilde{x}) + \tilde{D}\big(\mathbb{E}[Y \mid X = \tilde{x}]\big) \geq \mathbb{E}[Y \mid X = x]' \vec{u}(\tilde{x}) + \tilde{D}\big(\mathbb{E}[Y \mid X = x]\big).$$

Since $\tilde{D} \in \mathcal{D}$, it is finite at the referenced points because of optimality. The inequalities of the lemma follow from rearranging these inequalities. *Q.E.D.*

One feature captured in Lemma S.7.1 is the *monotonicity* condition

$$\big(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}]\big)'\big(\vec{u}(x) - \vec{u}(\tilde{x})\big) \geq 0. \tag{S.7}$$

This resembles the law of compensated demand if we relate $\mathbb{E}[Y \mid X = x]$ to Hicksian demand and $-\vec{u}(x)$ to the price vector. For further illustration of (S.7), suppose that $\vec{u}(x)$ and $\vec{u}(\tilde{x})$ only differ with respect to their first component. Then (S.7) becomes

$$\big(\mathbb{E}[Y_1 \mid X = x] - \mathbb{E}[Y_1 \mid X = \tilde{x}]\big)\big(u_1(x_1) - u_1(\tilde{x}_1)\big) \geq 0,$$

which states that the conditional expectation of Y_1 is weakly increasing in u_1.

We now use Lemma S.7.1 to remove the nuisance function $\tilde{D}$. To that end, let $x^0, \dots, x^{M-1}, x^M = x^0$ be a cycle of points in $\operatorname{supp}(X)$. By repeated application of Lemma S.7.1, we obtain

$$\begin{aligned}
&\sum_{m=0}^{M-1} \big(\mathbb{E}\big[Y \mid X = x^m\big] - \mathbb{E}\big[Y \mid X = x^{m+1}\big]\big)'\vec{u}\big(x^m\big) \\
&\quad \geq \sum_{m=0}^{M-1} \tilde{D}\big(\mathbb{E}\big[Y \mid X = x^{m+1}\big]\big) - \tilde{D}\big(\mathbb{E}\big[Y \mid X = x^m\big]\big) \\
&\quad = 0.
\end{aligned} \tag{S.8}$$

By summing up over a cycle, we "sum out" the unknown function $\tilde{D}$. An alternative way to state the inequalities obtained in this way is as follows. Suppose that $\{x^m\}_{m=0}^{M-1} \subseteq \operatorname{supp}(X)$. Then, for every permutation π of $\{0, \dots, M\}$, we have

$$\sum_{m=0}^{M-1} \mathbb{E}\big[Y \mid X = x^m\big]'\vec{u}\big(x^m\big) \geq \sum_{m=0}^{M-1} \mathbb{E}\big[Y \mid X = x^{\pi(m)}\big]'\vec{u}\big(x^m\big).$$

This inequality highlights the connection to optimizing behavior. Intuitively, no permutation can improve the "match" between choices (= conditional expectations) and payoffs (= marginal utility shifters). We now show that inequalities such as (S.8) capture the complete restrictions of the model for $\vec{u}$.

THEOREM S.7.1—Sharp Characterization of $\mathcal{U}_{\text{ID}}$: *Let $\vec{u} \in \mathcal{U}$. The following are equivalent*:

(i) *There is a function $\tilde{D} : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ such that*

$$\forall x \in \operatorname{supp}(X), \quad \mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k(x_k) + \tilde{D}(y).$$

(ii) $\vec{u} \in \mathcal{U}_{\text{ID}}$, *that is, there is a function $\tilde{D} \in \mathcal{D}$ such that*

$$\forall x \in \operatorname{supp}(X), \quad \mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k(x_k) + \tilde{D}(y).$$

(iii) *There exist finite numbers* $\{D_x\}_{x\in\text{supp}(X)}$ *such that, for every* $x, \tilde{x} \in \text{supp}(X)$,

$$\big(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}]\big)'\vec{u}(x) \geq D_{\tilde{x}} - D_x.$$

(iv) *For every integer M and cycle of points* $x^0, \dots, x^{M-1}, x^M = x^0$ *each in* $\text{supp}(X)$,

$$\sum_{m=0}^{M-1}\big(\mathbb{E}\big[Y \mid X = x^m\big] - \mathbb{E}\big[Y \mid X = x^{m+1}\big]\big)'\vec{u}\big(x^m\big) \geq 0.$$

PROOF: We shall show (i) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (ii) $\Longrightarrow$ (i).

By relating D_x with $\tilde{D}(\mathbb{E}[Y \mid X = x])$, the previous discussion shows (i) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv). Note that while $\tilde{D}$ may take on value $-\infty$ over some points, $\tilde{D}(\mathbb{E}[Y \mid X = x]) > -\infty$ for each $x \in \text{supp}(X)$. This is because $\tilde{D} \in \mathcal{D}$, $\vec{u}(x)$ is finite, and $\mathbb{E}[Y \mid X = x]$ a maximizer. This is why the numbers in (iii) are finite.

The implication (iv) $\Longrightarrow$ (ii) follows from Rockafellar (1970, Theorem 24.8), so we provide only a sketch of this implication. Let $\Gamma = \mathbb{R}^K \times \mathbb{R}^K$. Let $S = \{(\mathbb{E}[Y \mid X = x], \vec{u}(x))\}_{x\in\text{supp}(X)}$, so we have $S \subseteq \Gamma$. The set S is contained in the graph of a cyclically monotone multi-valued mapping (see Rockafellar (1970), which generalizes Definition S.7.1). By the constructive extension result of Rockafellar (1970, Theorem 24.8), we have $\vec{u}(x) \in \partial f(\mathbb{E}[Y \mid X = x])$, where f is a lower semi-continuous, convex function that never attains $-\infty$ and that is finite at some point. By Lemma S.1.1 and the fact that $\mathbb{E}[Y \mid X = x] \in \overline{B}$ for $x \in \text{supp}(X)$, we have

$$\mathbb{E}[Y \mid X = x]'\vec{u}(x) - f\big(\mathbb{E}[Y \mid X = x]\big) = \sup_{y\in\overline{B}}\big\{y'\vec{u}(x) - f(y)\big\}.$$

By letting $\tilde{D} = -f$, we have (ii).

Obviously, (ii) $\Longrightarrow$ (i). *Q.E.D.*

This result is closely related to results in Brown and Calsamiglia (2007) and Chambers and Echenique (2009). Related results that simultaneously vary budgets are established in McFadden and Fosgerau (2012).

The fact that (i) and (ii) are equivalent means that if we assume $\overline{D}$ is concave, we obtain no additional identifying power for $\vec{u}$. Moreover, it is not possible to separately test whether $\overline{D}$ is concave aside from testing the entire model. This insight is fairly well-known in other settings (Afriat (1967), Varian (1982)).

Part (iii) is helpful for computational reasons such as checking whether a particular point is in the identified set. Note that we need not worry about forcing D_x and $D_{\tilde{x}}$ to agree whenever $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y \mid X = \tilde{x}]$, since (iii) implies $D_x = D_{\tilde{x}}$ by double inequalities.

If the parameter space $\mathcal{U}$ contains constant functions, these functions will always be in $\mathcal{U}_{\text{ID}}$. This can easily be seen from (iv). We refer to (iv) as the *cyclic monotonicity* inequalities in light of the following definition.

DEFINITION S.7.1—Cyclic Monotonicity: $f : \mathbb{R}^\ell \to \mathbb{R}^\ell$ is said to be *cyclically monotone* if, for every integer M and cycle $x^0, x^1, \dots, x^{M-1}, x^M = x^0$ of points each in $\mathbb{R}^\ell$,

$$\sum_{m=0}^{M-1}\big(f\big(x^m\big) - f\big(x^{m+1}\big)\big)'x^m \geq 0.$$

Cyclic monotonicity has been used in the econometrics literature by McFadden and Fosgerau (2012) and Shi, Shum, and Song (2018).

REMARK S.7.1—Single-Dimensional Case: When $K = 1$, it can be shown that Theorem S.7.1(iv) is equivalent to the condition that, for every $x, \tilde{x} \in \mathrm{supp}(X)$,

$$\big(\mathbb{E}[Y_1 \mid X = x] - \mathbb{E}[Y_1 \mid X = \tilde{x}]\big)\big(u_1(x_1) - u_1(\tilde{x})\big) \geq 0.$$ [6]

Note that when $K = 1$, u_1 is the only utility index. This shows that when $K = 1$, the most we can say about u_1 is that it must be consistent with the monotonicity statement:

$$\mathbb{E}[Y_1 \mid X = x] > \mathbb{E}[Y_1 \mid X = \tilde{x}] \quad \Longrightarrow \quad u_1(x) \geq u_1(\tilde{x}).$$ [7]

This is purely ordinal information, and so point identification is impossible for many choices of the parameter space. If $\mathcal{U}$ is unrestricted, then in the single-dimensional case whenever $\tilde{u} \in \mathcal{U}_{\mathrm{ID}}$, we also have $g(\tilde{u}) \in \mathcal{U}_{\mathrm{ID}}$ for any strictly increasing function g. Even if $\mathcal{U}$ is restricted to a class of differentiable functions with a location/scale normalization, $\mathcal{U}_{\mathrm{ID}}$ may not be a singleton.

REFERENCES

AFRIAT, S. N. (1967): "The Construction of Utility Functions From Expenditure Data," *International Economic Review*, 8 (1), 67–77. [11]

BERRY, S., A. GANDHI, AND P. HAILE (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81 (5), 2087–2111. [5]

BROWN, D. J., AND C. CALSAMIGLIA (2007): "The Nonparametric Approach to Applied Welfare Analysis," *Economic Theory*, 31 (1), 183–188. [11]

CHAMBERS, C. P., AND F. ECHENIQUE (2009): "Profit Maximization and Supermodular Technology," *Economic Theory*, 40 (2), 173–183. [11]

FOX, J. T. (2007): "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices," *The RAND Journal of Economics*, 38 (4), 1002–1019. [3]

GOEREE, J. K., C. A. HOLT, AND T. R. PALFREY (2005): "Regular Quantal Response Equilibrium," *Experimental Economics*, 8 (4), 347–367. [3]

HAN, A. K. (1987): "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35 (2), 303–316. [12]

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3 (3), 205–228. [3,4]

MATZKIN, R. L. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58 (1), 137–168. [3]

MCFADDEN, D. L., AND M. FOSGERAU (2012): "A Theory of the Perturbed Consumer With General Budgets," Technical report, National Bureau of Economic Research. [11,12]

ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press. [1,11,12]

SHI, X., M. SHUM, AND W. SONG (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86 (2), 737–761. [12]

STINCHCOMBE, M. B., AND H. WHITE (1992): "Some Measurability Results for Extrema of Random Functions Over Random Sets," *The Review of Economic Studies*, 59 (3), 495–514. [2]

VARIAN, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50 (4), 945–973. [11]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 16 November, 2017; final version accepted 11 January, 2019; available online 13 February, 2019.

[6] See Rockafellar (1970, p. 240).

[7] When u_1 is assumed linear, these restrictions are implied by but do not generally imply the restrictions of the generalized regression model of Han (1987).